

本周周报（2013.12.16-2013.12.22）

郭方舟

本周工作

1 空气污染数据可视化

本周主要是在复习托福和完成作业，因此在这个任务上投入的时间不多。

CCA:

本周主要看了一下 CCA 算法，上一次在尝试 CCA 之后觉得这个算法对项目的用处不大，而周六顾教授说 CCA 会有用，因此又找了一些关于 CCA 的资料。

CCA 是针对两组变量 X, Y 进行的。其目标是找到一对向量 (u, v) ，使得 X 经过 u 变换后得到的结果与 Y 经过 v 变换后得到的结果的相关性最大。在找到第一对向量之后，再找出与第一对向量无关的第二对向量使变换后结果的相关性最大，以此类推。因此，CCA 其实可以算是一种投影算法。

在进行 CCA 之前，首先要计算 X, Y 之间的相关性。如果 X, Y 之间没有相关性，那么进行 CCA 就没有意义了。

当一对相关向量对应的相关系数足够大时，可以用一组变量的值预测另一组变量的值。

在看了一些相关的讲义和论文之后，我感觉使用 CCA 来找气象数据中各组变量的相关性有一些奇怪，因为在进行计算典型相关系数之前，首先要计算每组变量之间的相关性，这样就有点绕圈子了。我打算下周找一些使用了 CCA 方法的论文来看一下，看看别人是怎么用 CCA 这个方法的。

转移熵:

转移熵方面的工作这周主要是斐然师兄在做，目前的方法是在进行概率统计之后，对概率矩阵做了一个 kde，这种方法做出来的 TE 结果比原来的结果要好一些。同时，原始数据的分割上也进行了一些修改，原先是分为十个等级，现在只分四个等级。斐然师兄的意思是 TE 的工作基本先做到这里，接下来主要研究 CCA，时序回归分析和 VAR。

2 课程作业

完成网格计算的作业，与小马哥一起写了两个基于 SPARK 平台的投影算法，分别是 LMDS 和 LAMP。

12 月 21 号参加了托福考试。

下周工作

1. 空气污染数据的可视化

找 CCA 相关的例子进行学习和实现。

找时序回归分析的论文和例子进行学习和实现